View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Volume 240, number 1,2, 33–40 FEB 06494 November 1988

Conserved and variable elements in RNA genomes of potexviruses

K.G. Skryabin, S.Yu. Morozov, A.S. Kraev, M.N. Rozanov, B.K. Chernov, L.I. Lukasheva and J.G. Atabekov

Department of Virology, Moscow State University, Moscow 119899 and Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow 117984, USSR

Received 22 September 1988

The nucleotide sequences of genomic RNAs and predicted amino acid sequences of two strains of potato virus X and white clover mosaic potexvirus were compared to each other, and the proteins of different plus-RNA-containing plant viruses. The predicted non-virion proteins of potexviruses have direct sequence homology and common structural peculiarities with those of several 'Sindbis-like' plant viruses. The most conserved amino acid sequences were found to be located in the polypeptide encoded by the long 5′-proximal open reading frame (ORF1). The putative polypeptide encoded by the ORF2 starting beyond the ORF1 stop codon is clearly related to the presumptive NTP-binding domain of the ORF1-coded polypeptide. These results suggest possible functions for all of the potexvirus proteins and also indicate that potexviruses have a genome organization which is considerably different from that of other plant viruses.

Genome organization; Potexvirus; Sequence homology; Plant virus

1. INTRODUCTION

The potexviruses are the filamentous plant viruses consisting of multiple copies of a single coat protein of 20–28 kDa [1] and a single-stranded infectious RNA of 2.1–2.3 MDa [2] which is 3′-polyadenylated [3–5], contains a 5′-terminal cap structure [6], and is expressed in its 3′-terminal part through the production of the 3′-coterminal subgenomic mRNAs [5,7–9]. The primary structures of the genomes of white clover mosaic virus (WClMV), Russian strain of potato virus X (PVX) and PVX strain X3 have been completed recently [PVX genome length, 6435 bases; WClMV, 5845 bases, excluding poly(A)] [10–12]. The nucleotide sequences of the extended 3′-terminal regions have been reported for potato aucuba mosaic virus (PoAMV) RNA [13] and papaya mosaic virus (PMV) RNA [14]. Potential cistrons capable of encoding five polypeptides are present in the potexvirus genomes, although no such proteins, excepting the capsid protein, have yet been identified in vivo. The order of the cistrons in the PVX genome has been determined as follows: 5′ – ORF1 (165 kDa protein gene) – ORF2 (25 kDa protein gene) – ORF3 (12 kDa protein gene) – ORF4 (8 kDa protein gene) – ORF5 (coat protein gene) – 3′ [11,12].

Here, we carried out intra-group and inter-group comparisons of the nucleotide sequences and predicted potexvirus amino acid sequences. These comparisons have enabled us to make suggestions for the functions of RNA sequences and some of the potexviral putative proteins.

Correspondence address: J.G. Atabekov, Dept of Virology, Moscow State University, Moscow 119899, USSR

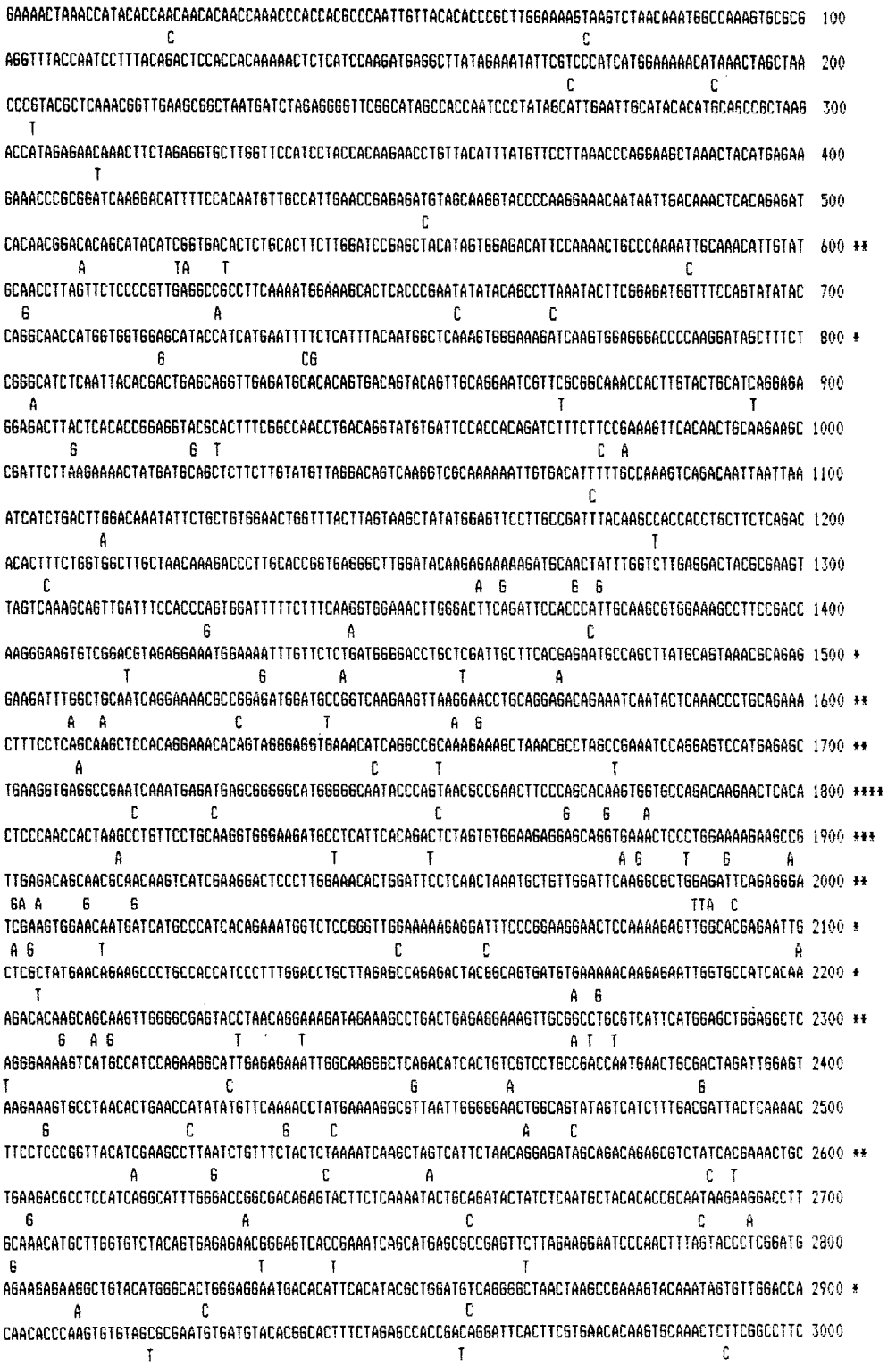
The nucleotide sequences presented here have been submitted to the EMBL/GenBank database under the accession no. Y00974

2. EXPERIMENTAL

The amino acid sequences of the PVX and WClMV proteins were predicted from the nucleotide sequences of the viral RNAs [10–12].

For initial homology search, dot matrix comparisons were performed using the computer program DIAGON [15]. For construction of the dot matrix graph between the PVX and

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50 © 1988 Federation of European Biochemical Societies



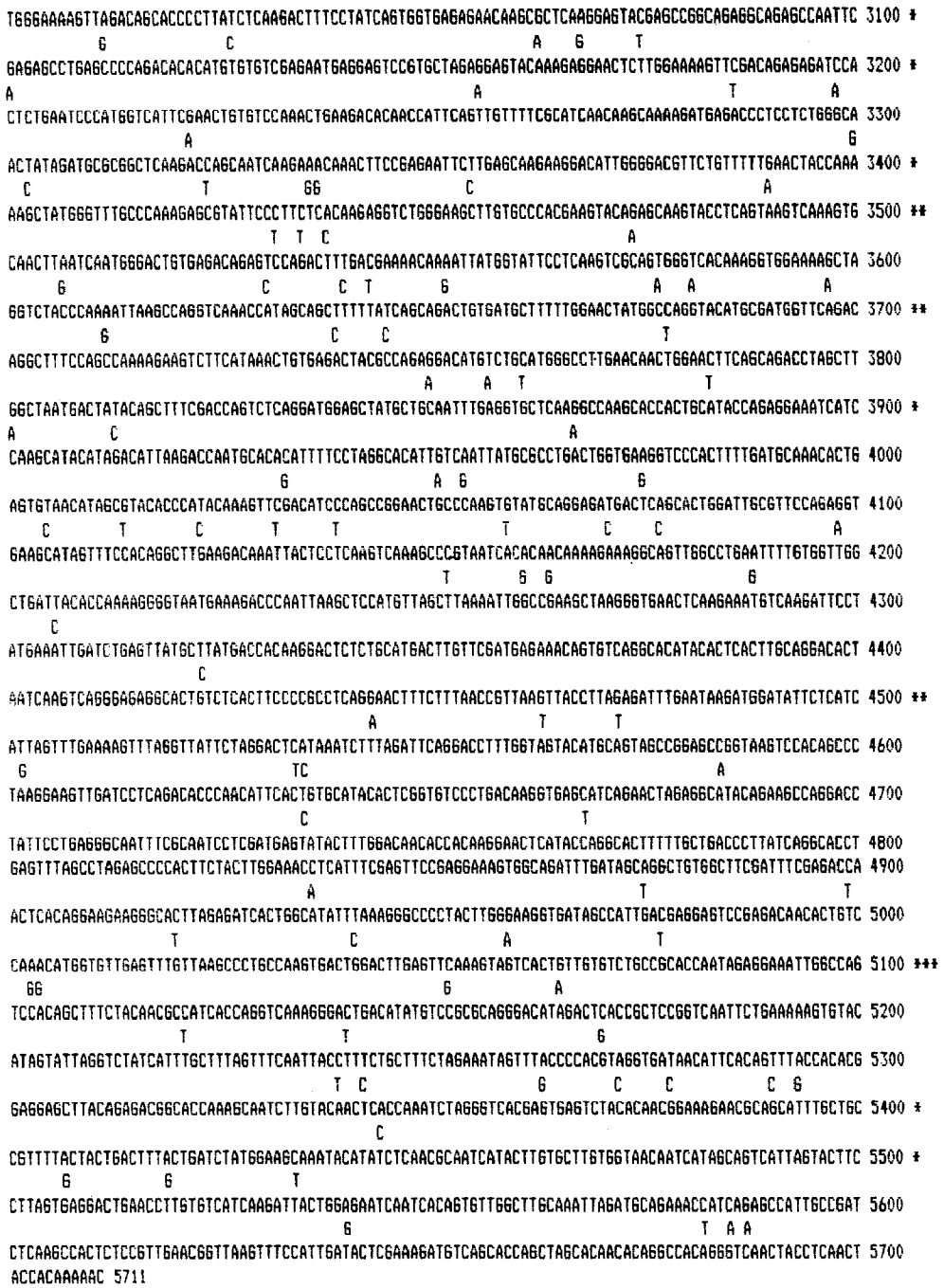


Fig.1. Alignment of the nucleotide sequences of two PVX strains. Nucleotide sequence of PVX (Russian strain) RNA is shown in the DNA form with nucleotide numbers on the right. Base exchanges for PVX (strain X3) are indicated below the numbered sequence. The total number of amino acid substitutions between two PVX strains is marked by asterisks on the right of each line.

WClMV high molecular mass polypeptides, all possible 21-amino-acid blocks from each sequence were compared. Those comparisons producing a score greater than 237 were represented by a single point on the graph. The detailed alignments were made by visual inspection of the homologous regions revealed using dot plots. Sequences were analysed using a WICAT-S computer.

3. RESULTS AND DISCUSSION

3.1. *Nucleotide sequences of the potexviral genomes: intra-group, inter-strain and intra-genomic comparisons*

The complete nucleotide sequences of the genomic RNA for PVX (Russian strain) [11] and PVX (strain X3) [12] have recently been determined. Comparison of the RNA sequences coding non-virion proteins in Russian strain with those reported for strain X3 is shown in fig.1. This comparison permits us to establish base substitution rate with high accuracy because in both cases the nucleotide sequences were determined by use of dideoxy sequencing. The untranslated 5′-end leader region displays a low degree of base substitutions (2.4%) when compared with the ORF1 region (4.2%), triple block of ORFs 2–4 (3%) and the intercistronic region between ORF1 and ORF2 (6.7%); the average substitution rate being 3.9%. Significant sequence conservation was found also between the 5′-terminal leader regions of different potexviral RNAs (PVX, WClMV and PMV) [10]. This region has been previously identified as the site for assembly nucleation [4].

An additional region of sequence conservation between the potexviral genomes was found in the area upstream from the coat protein gene [10,13,16]. The consensus sequence revealed (fig.2A) very likely represents a promoter for initiating the synthesis of subgenomic RNAs at the minus-strand RNA, as shown for some other plant viral plus-RNA genomes [17]. Among the

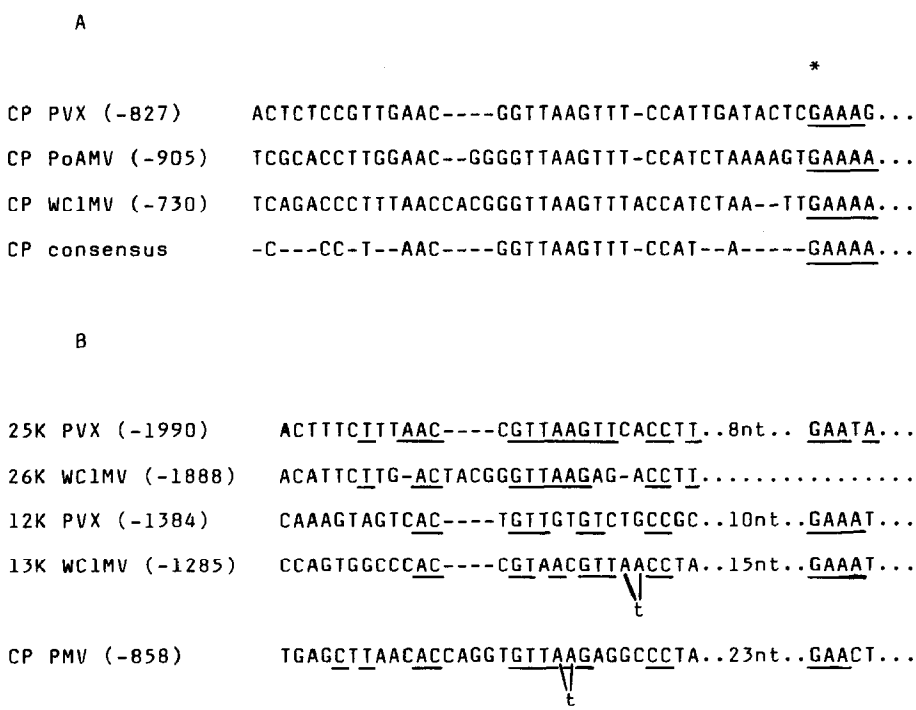


Fig.2. Comparison of the putative potexvirus subgenomic promoter sequences. (A) The sequences of the three coat protein mRNA promoters. Asterisk indicates the initiation point of the PVX coat protein mRNA (Morozov, S.Yu., unpublished). The motifs typical for the 5′-terminal nucleotide sequence of potexvirus genomes are underlined. Gaps (−) have been introduced for maximum alignment. The numbering system used shows positions with respect to the 3′-terminal poly(A) sequence. (B) Sequence elements of the five putative potexvirus promoters. The features characterizing coat protein mRNA promoter consensus are underlined. Numerals show the numbers of nucleotides which could not be aligned. Two residues are shown under the sequences of 13 kDa WClMV and CP PMV promoters for maximum alignment.

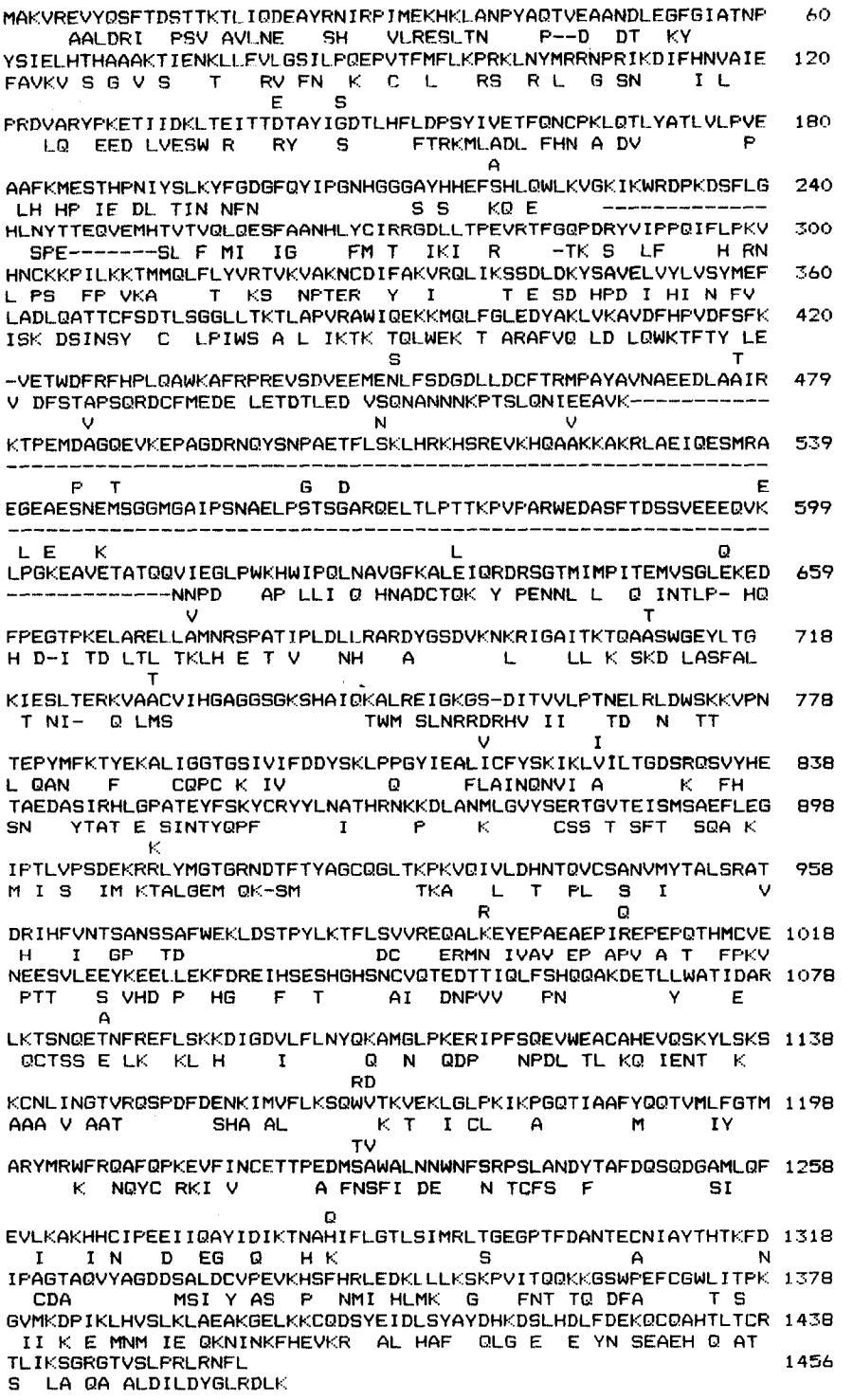


Fig.3. Alignment of the ORF1-coded proteins of the potexviruses. Amino acid sequence of the PVX (Russian strain) 165 kDa protein is numbered on the right. Amino acid exchanges for WClMV and PVX (strain X3) are indicated below and above the numbered sequence, respectively. Gaps (−) are introduced in the sequences for maximum alignment.

predicted promoter elements of potexviruses we also map the minor subgenomic RNA promoters upstream of the ORF2 and ORF3 of PVX and WClMV (fig.2B). It is significant that the expected size of the RNAs transcribed using the latter putative promoters agrees well with the length of the 2.1 kb and 1.4 kb subgenomic RNAs detected in the potexvirus-infected plants [5,9].

3.2. *The 165 kDa protein: inter-strain, intra-group and inter-group comparisons*

The occurrence of mutations between two PVX strains is not randomly distributed along the RNA sequence (fig.1). In the sequence of the 165 kDa protein the number of amino acid changes is 2.2% (fig.3). The changes observed in the 165 kDa protein sequence derived from Russian strain and from strain X3 indicate one domain especially susceptible to substitutions: the region 470–615 contains 8.3% of amino acid substitutions (figs 1,3). It is very interesting that this region is deleted from the WClMV 147 kDa protein (fig.3). It can be speculated that this region is not the functionally essential part of the PVX 165 kDa protein in contrast to regions 1–230 and 617–1450 which could be aligned with the WClMV 147 kDa protein as 52–53% of compared amino acids are identical (fig.3).

There are good reasons to suggest that the region 617–1450 of the PVX 165 kDa protein is involved in the replication of viral RNA. This region was shown to contain two domains of homology with two RNA-replicating proteins of brome mosaic virus and other Sindbis-like viruses [10,11,18,19]. The first domain was homologous to the hypothetical NTP-binding domain [20] (fig.4). The conserved domain including invariant residues DXXXXD, T/SXXXTXXNT/S (where X may be any amino acid), and the GDD sequence flanked as a rule by uncharged residues found in all RNA-dependent RNA polymerases of positive-strand RNA viruses [10,11,21] is also present in the 165 kDa PVX protein (fig.5). This second domain was also significantly homologous to the analogous protein sequences of Sindbis-like viruses. The most considerable homology was found unexpectedly between the filamentous potexviruses and spherical tymoviruses (fig.5).

3.3. *The 25 kDa protein: intra-viral and inter-group comparisons*

The 25 kDa protein of PVX and the 26 kDa protein of WClMV display a significant relation with the conserved NTP-binding domains of the proteins of Sindbis-like viruses [20] (fig.4). A quite unexpected feature of the potexviral genomes is

```
1.PVX     25K  25_VVHAVAGAGKSTALRKLI_34_FAILDEYTLDN_ 6_QAIFADPYQAPE_
2.WClMV   26K  24_VVHAIAGSGKSTVIRKIL_34_LDILDEYGQLP_ 8_EFIFTDPYQAPT_
3.BSMV    58K 266_IISGVPGSGKSTIVRTLL_38_LLIIDEYTLAE_11_VLLVGDVAQGKA_
4.BNYVV   43K 120_IVLGAPGVGKSTSIKNLL_46_TMLVDEVTRVH_11_VICFGDPAQGLN_
5.PVX    165K 731_VIHGAGGSGKSHAIQKAL_47_IVIFDDYSKLP_16_VILTGDSRQSVY_
6.WClMV  147K 566_VIHGAGGSGKSHAIQTWM_48_IIVFDDYSKLP_16_AILTGDSKQSFH_
7.BNYVV  237K 892_YVKGGPGTGKSFLIRSLA_45_IIFVDEFTAID_11_IYLVGDEQQTGI_

"NTPase-helicase" V G AG GKS              VDE             GD  Q
  consensus       I A P    T              I                S
                  Y                       F

              1.  _10_TSFRVPRKV_47_FVKPCQVTGLEFKVVTVVS_10_AFYNAITRS_
              2.  _11_TTYRFGPNT_47_FFKVSDVIGYQWPTVTLYL_13_LLFIGLTRH_
              3.  _18_TTYRLGQET_62_CALAIDVQGKEFDSVTLFL_12_LRLVALSRH_
              4.  _18_ASRRFGKAT_67_SILYSDAHGQTYDVVTIIL_13_VRAVLLTRA_
              5.  _28_ATHRNKKDL_40_TFTYAGCQGLTKPKVQIVL_10_VMYTALSRA_
              6.  _28_ITHRNKPDL_40_SMTYAGCQGLTTKAVQILL_10_VIYTALSRA_
              7.  _25_MNFRNPVHD_72_KTTVRANQGSTYDNVVLPV_12_LNLVALSRH_

"NTPase-helicase"      R           T   SQG E   V          VALSR
  consensus                        L   AK  S   A          TLVT
                                       VH  T              GM
                                       NA  R              SI
```

Fig.4. Alignment of amino acid sequences of the 25 kDa protein of PVX, 26 kDa protein of WClMV, 165 kDa protein of PVX and 147 kDa protein of WClMV with NTP-motif containing proteins of barley stripe mosaic virus (BSMV) and beet necrotic yellow vein virus (BNYVV). Helicase consensus according to [24] is shown. Viral protein sequences are from sources cited in [24].

```
 1.PVX    165K 1238_SLANDYTAFDQSQDGAMLQFEVL_23_IFLGTLSIMRLTGEGPTFDANTECNIAYTHTKF_
 2.WClMV  147K 1073_CFSNDFTAFDQSQDGSILQFEVI_23_IFLGTLSIMRLSGEGPTFDANTEANIAYTHTKF_
 3.TYMV   206K 1574_KIANDYTAFDQSQHGESVVLEAL_23_TQFGPLTCMRLTGEPGTYDDNTDYNLAVIYSQY_
 4.BNYVV  237K 1837_NGVIDAAACDSGQGVFTQLIERH_27_YVRAHMSYVKTSGDVTTFIGNTVIIAACLASML_
 5.TMV    183K 1382_VLELDISKYDKSQNEFHCAVEYE_29_GIKTCIWYQRKSGDVTTFIGNTVIIAACLASML_
 6.TRV    194K 1451_FVEIDMSKFDKSANRFHLQLQLE_29_GMMAHIWYQQKSGDADTYNANSDRTLCALLSEL_
 7.BSMV    74K  404_ALEIDFSKFDKSKTGLHIKAVIG_29_GLEAYLLYQQKSGNCDTYGSNTWSAALALLDCL_
 8.CMV     94K  514_CLEIDLSKFDKSQGEFHLMIQEH_29_GVGMPISFQRRTGDAFTYFGNTIVTMAEFAWCY_
 9.BMV     94K  465_FLEADLSKFDKSQGELHLEFQRE_29_KVGMSVSFQRRTGDAFTYFGNTLVTMAMIAYAS_
10.AlMV    90K  530_FKEIDFSKFDKSQNELHHLIQER_29_GVFFNVDFQRRTGDALTYLGNTIVTLACLCHVY_
11.SNBV  nsP4  489_VLETDIASFDKSQDDAMALTGLM_29_GTRFKFGAMMKSGMFLTLFVNTVLNVVIASRVL_

12.YFV   ns5   532_FYADDTAGWDTRITEADLDDEQE_38_MDVISRRDQRGSGQVVTYALNTITNLKVQLIRM_
13.IBV    F2   573_LMGWDYPKCDRAMPNLLRIAASL_33_GGIYVKPGGTSSGDATTAYANSVFNIIQATSAN_
14.BBV   102K  584_VIETDFSNLDGRVSSWMQRNIAQ_32_GFRIEPGVGVKSGSSTTTPHNTQYNGCVEFTAL_
15.CarMV  86K  470_AIGFDMSRFDQHVSVAALEFEHS_31_MLRYTKEGCRMSGDMNTALGNCLLACLITKHLM_
16.SBMV  103K  696_AAEADISGFDWSVQDWELWADVE_33_LLQQELPGIMKSGSYCTSSTNSRIRCLMAELIG_
17.IBDV   90K  411_WYSIDLEKGEANCTRQHMQAAMY_38_MNLQIKSYGQGSGNAATFINNHLLSTLVLDQWN_
18.MS2    pol  254_LATIDLSSASDSISDRLVWSMLP_18_GETIRWELFSTMGNGFTFELESMIFWAIVKATQ_

"polymerase"          D    D                                SG   T   NT
 consensus                                                   T         S

                1._ 7_QVYAGDDSALDC_31_PEFCGWLITPK_ 2_MKDPIKLHVSLKLAEA_
                2._ 7_QVYAGDDMSIDY_31_AEFCGWTISPK_ 2_IKKPEKMNMSIELQKN_
                3._ 6_IMVSGDDSLIDH_26_PLFCGYYVGPA_ 2_IRNPLALFCKLMIAVD_
                4._ 7_MAMKGDDGFKRQ_29_ITFCGYALSNG_ 0_HLFPSV-SRKLTKIAA_
                5._ 7_GAFCGDDSLLYF_28_GYFCGRYVIHH_ 8_YYDPLKLISKLGAKHI_
                6._ 7_VTYGGDDSLIAF_28_PMFCGKFLLKT_ 6_VPDPVKVLTKLGKKSI_
                7._ 7_CVFGGDDSLILF_28_PAFCGKFLLCI_ 6_VPDAAKFITKLGRTDV_
                8._ 7_LLFSGDDSLAFS_26_PYICSKFYSLM_ 6_QSPTIREIQRLGTKKI_
                9._ 7_AIFSGDDSLIIS_26_PYVCSKFLVET_ 7_VPDPLREIQRLAKRKI_
               10._ 9_VVASGDDSLIGT_27_PFICSKFLITM_11_IPNPLKLLIRLGSKKV_
               11._ 9_AAFIGDDNIIHG_30_PYFCGGFILQD_ 8_VADPLKRLFKLGKPLP_

               12._37_MAVSGDDCVVRP_36_VPFCSHHFHEL_ 7_IVVPCREQDELIGRGR_
               13._51_LMILSDDGVVCY_41_HEFCSQHTMLV_10_YPDPSRILGACVFVDD_
               14._16_GPKCGDDGLSRA_25_IGLCFLSRVFV_ 8_IQDPLRTLRKLHLTTR_
               15._ 5_LINNGDDCVLIC_34_IRFCQMAPVFD_ 6_VRDPLVSMSKDSHSLV_
               16._ 3_CIAMGDDSVEGF_31_VEFCSHVIKKR_ 4_LTSWPKTLYRFLSTPR_
               17._24_IERSIDDIRGKL_50_RLFCSAAYPKG_10_GIEQAYKVVRYEALRL_
               18._ 8_IGIYGDDIICPS_28_RESCGAHFYRG_ 3_KPFYIKKPVDNLFALM_

  "polymerase"          GDD            FCG             P
   consensus                             S
```

Fig.5. Comparison of the conserved regions of the putative RNA-dependent RNA polymerases. Viral protein sequences are from sources cited in [21]. Turnip yellow mosaic virus (TYMV) sequence has been published in [22]. The sequence of the protein of birnavirus (infectious bursal disease virus, IBDV) has been published in [23].

that two different proteins (25 and 165 kDa in PVX, 26 and 147 kDa in WClMV) contain distantly homologous domains related to bacterial DNA helicases [24] (fig.4). Gorbalenya et al. [20] have pointed to the duplicate NTP-binding domains in the multipartite genomes of furoviruses and hordeiviruses.

3.4. *The 12 and 8 kDa proteins: intra-group and inter-group comparisons*

The potexvirus 12 and 8 kDa proteins encoded in overlapping ORFs 3 and 4 contain blocks of uncharged amino acids which resemble the essential features of the membrane-spanning segments [10,16]. Morozov et al. [16] have revealed the principal similarity in the organization of these ORFs between potexviruses, hordeiviruses and furoviruses. The resemblance between the genomes of the latter viruses is more prominent, since blocks of similar overlapping ORFs also include the NTP-binding protein cistron [10,12,25].

It was demonstrated [26] that the cell-to-cell movement of TMV mutant, Ls1 (*ts* mutant in transport function), can be complemented by PVX in plants with mixed infections. Consequently, the PVX genome is capable of coding for a protein(s)

which is a functional analog of the TMV 30 kDa transport protein. However, the proteins structurally homologous to the TMV 30 kDa protein do not seem to be encoded by the genomes of all viruses which are known to contain the triple block of the overlapping ORFs coding for the NTP-binding protein and two membrane-bound proteins. On the other hand, the genomes of viruses coding for the structural homologs of the TMV 30 kDa protein do not contain this triple block of genes. One can speculate that one or both of the potential membrane-bound proteins coded by potexviruses, hordeiviruses and furoviruses are involved in transport function.

REFERENCES

[1] Purcifull, D.E. and Edwardson, J.R. (1981) in: Handbook of Plant Virus Infections and Comparative Diagnosis (Kurstak, E. ed.) pp.627–693, Elsevier/North-Holland, Amsterdam, New York.

[2] Koenig, R. (1971) J. Gen. Virol. 10, 111–114.

[3] Morozov, S.Yu., Gorbulev, V.G., Novikov, V.K., Agranovsky, A.A., Kozlov, Yu.V., Atabekov, J.G. and Baev, A.A. (1981) Dokl. Akad. Nauk SSSR 259, 723–725.

[4] Abouhaidar, M.G. (1983) Can. J. Microbiol. 29, 151–156.

[5] Guilford, P.J. and Forster, R.L.S. (1986) J. Gen. Virol. 67, 83–90.

[6] Sonenberg, N., Shatkin, A.J., Ricciardi, R.P., Rubin, H. and Goodman, R.M. (1978) Nucleic Acids Res. 5, 2501–2512.

[7] Short, M.N. and Davies, J.W. (1983) Biosci. Rep. 3, 837–846.

[8] Bendena, W.G., Bancroft, J.B. and Mackie, G.A. (1987) Virology 157, 276–284.

[9] Dolja, V.V., Grama, D.P., Morozov, S.Yu. and Atabekov, J.G. (1987) FEBS Lett. 214, 308–312.

[10] Forster, R.L.S., Bevan, M.W., Harbison, S.-A. and Gardner, R.C. (1988) Nucleic Acids Res. 16, 291–303.

[11] Kraev, A.S., Morozov, S.Yu., Lukasheva, L.I., Rozanov, M.N., Chernov, B.K., Simonova, M.L., Golova, Yu.B., Belzhelarskaya, S.N., Pozmogova, G.E., Skryabin, K.G. and Atabekov, J.G. (1988) Dokl. Akad. Nauk SSSR 300, 711–716.

[12] Huisman, M.J., Linthorst, H.J.M., Bol, J.F. and Cornelissen, B.J.C. (1988) J. Gen. Virol. 69, 1789–1798.

[13] Bundin, V.S., Vishnyakova, O.A., Zakharyev, V.M., Morozov, S.Yu., Atabekov, J.G. and Skryabin, K.G. (1986) Dokl. Akad. Nauk SSSR 290, 728–733.

[14] Abouhaidar, M.G. (1988) J. Gen. Virol. 69, 219–226.

[15] Staden, K. (1982) Nucleic Acids Res. 10, 2951–2961.

[16] Morozov, S.Yu., Lukasheva, L.I., Chernov, B.K., Skryabin, K.G. and Atabekov, J.G. (1987) FEBS Lett. 213, 438–442.

[17] French, R. and Ahlquist, P. (1988) J. Virol. 62, 2411–2420.

[18] Goldbach, R.W. (1987) Microbiol. Sci. 4, 197–202.

[19] Quadt, R., Verbeek, H.J.M. and Jaspars, E.M.J. (1988) Virology 165, 256–261.

[20] Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1988) J. Mol. Evol., in press.

[21] Koonin, E.V., Chumakov, K.M., Yushmanov, S.V. and Gorbalenya, A.E. (1988) Mol. Genet. 3, 16–19.

[22] Morch, M.-D., Boyer, J.-C. and Haenni, A.L. (1988) Nucleic Acids Res. 16, 6157–6190.

[23] Morgan, M.M., Macreadie, I.G., Harley, V.R., Hudson, P.J. and Azad, A.A. (1988) Virology 163, 240–242.

[24] Hodgman, T.C. (1988) Nature 333, 22–23.

[25] Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K. and Jonard, G. (1986) J. Gen. Virol. 67, 1689–1700.

[26] Taliansky, M.E., Malyshenko, S.I., Pshennikova, E.S. and Atabekov, J.G. (1982) Virology 122, 327–331.